

Cosmic Origins Science Enabled by the WFIRST-AFTA Archives

Sally Heap (GSFC) and the SAG8 Team

Carter Chamberlain / Va Tech

Dennis Ebbets / DCEbbets@Gmail.com

Caryl Gronwall/ Penn State

Mario Juric / LSST

Anton Koekemoer / STScI

Jeff Kruk / NASA GSFC

Tom McGlynn / NASA GSFC

Neill Reid / STScI

Chris Stockdale / Marquette U.

Harry Teplitz / IPAC

16 July 2015

TABLE OF CONTENTS OF SHORTER REPORT

ABSTRACT, P. 3

SECTION 1: INTRODUCTION, P. 4

SAG8 CHARTER, P. 4

A PREVIEW, P. 5

ORGANIZATION OF REPORT, P. 7

SECTION 2: GENERAL ASTROPHYSICS WITH WFIRST-AFTA, P. 8

[READERS ARE DIRECTED TO THE WFIRST SDT 2015 REPORT]

SECTION 3: WFIRST OBSERVATIONS, DATA PROCESSING, DATA PRODUCTS,
P. 27

INTRODUCTION, P. 28

HIGH-LATITUDE IMAGING SURVEY, P. 31

HIGH-LATITUDE SPECTROSCOPIC SURVEY, P. 33

IFU SURVEY OF SUPERNOVAE AND GALAXIES, P. 35

MICROLENSING SURVEY OF THE GALACTIC BULGE, P. 337

SPECIAL CALIBRATIONS, P. 39

SECTION 4: ACCESSING THE WFIRST ARCHIVES, P. 41

INTRODUCTION TO THE QUERY SYSTEM, P. 42

SAMPLE QUERIES OF THE WFIRST ARCHIVE, P. 44

ABSTRACT

The Cosmic Origins science theme embraces research into the origins, formation and evolution of the constituents of our universe, including galaxies, stars, and planetary systems. It reaches from the earliest observable epochs to today's modern universe. The 2010 Decadal Survey acknowledged these quests under its theme of Cosmic Dawn. Although many research approaches use detailed high-resolution observations of known individual objects/systems, other approaches use the observable systematics of large numbers of related objects, such as stellar color-magnitude diagrams and stellar or galaxy luminosity functions. In some cases, important objects are rare, and new members of a class need large surveys to identify them for additional study. As described in Section 2 of this report, the planned WFIRST imaging and spectroscopic surveys will produce huge volumes of new data on the contents of our Milky Way galaxy, the Local Group of galaxies, and the universe extending out to redshifts at the epoch of reionization and farther. Observations by Guest Observers will increase the volume and value of the data. As described in Section 3, we find that the surveys conducted by WFIRST will produce important information for cosmic-origins science *without modification to the observing program*. Furthermore, it is likely that most of the data processing needed for dark-energy science and exoplanet science will also serve for cosmic-origins purposes. Finally, as described in Section 4, we find that thoughtful design of the data archives, particularly the means by which scientists will identify, access, and extract information from the archives will be critical to the ultimate productivity of this flagship mission. A powerful approach to a modern user interface is a query system as found in the SDSS SkyServer, the Hubble Source Catalogue, and NASA'S WISE and Spitzer archives. Section 4 of this report gives examples of queries that return the most relevant information and can be used to discover complex relationships. As the design of the WFIRST hardware, software, ground system and archive matures in the future, these examples will help ensure a system of great value to a very wide scientific community.

SECTION 1: INTRODUCTION

Wide-Field Infrared Survey Telescope (WFIRST) was envisaged by the 2010 Astrophysics Decadal Survey Panel as an “observatory designed to settle essential questions in both exoplanet and dark energy research, and which will advance topics ranging from galaxy evolution to the study of objects within our own galaxy”. The transfer of the Astrophysics-Focused Telescope Assets (AFTA) to NASA in 2013 enabled a new, more powerful version of WFIRST, which addresses all three areas of NASA’s astrophysics program: Physics of the Cosmos (How does the universe work?) Exoplanet Exploration (Are we alone?), and Cosmic Origins (How did we get here?)

With the submission of 1-page science ideas, it became clear that WFIRST-AFTA observational data obtained for dark energy research or exoplanet demography also could, without modification, address Cosmic-Origins issues. The Final Report of the Science Definition Team (SDT) acknowledges the value of WFIRST-AFTA to Cosmic-Origins research in its Report (http://wfirst.gsfc.nasa.gov/science/sdt_public/WFIRST-AFTA_SDT_Report_150310_Final.pdf) by devoting two sections of its report to general astrophysics. These two sections are reprinted as Section 2 of this report.

In 2014, the Cosmic Origins Program Analysis Group (COPAG) established Science Analysis Group #8 (SAG8) to explore and analyze cosmic-origins science enabled by the WFIRST-AFTA archive. The SAG8 charter is given below:

SAG8 Charter

Cosmic Origins Science Enabled by the WFIRST-AFTA Data Archive

The Wide-Field Infrared Survey Telescope (WFIRST) is the highest priority large space mission recommended by the recent decadal survey in astronomy and astrophysics. It is designed to perform wide-field imaging and slitless spectroscopic surveys of the visible to near-infrared sky. The Astrophysics Focused Telescope Assets (AFTA) study design of the mission makes use of an existing 2.4m telescope to enhance light collecting and imaging performance. The main instrument is a wide-field multi-filter imager with infrared grism spectroscopy. It also features a small-field low-resolution integral field spectrograph. A coronagraph instrument was part of the study and has a primary science focus of direct imaging of gas-giant exoplanets and debris disks.

Achieving the full science potential of WFIRST-AFTA will require input from the astronomical community on how it intends to use the vast WFIRST-AFTA data archive for Cosmic Origins science. The infrared surveys and coronagraphic investigations will provide abundant opportunities for archival research. The high latitude wide-field infrared survey alone is expected to observe more than 500 million galaxies over a 2000 square degree area at a resolution of about 0.11 arc

seconds in four broad near-infrared passbands. An active Guest Observer program will further populate the archive with a multitude of datasets. A cross-section of archival Cosmic Origins science investigations would be valuable input in the formative stages of the mission for discussions of high-level science products, catalogs, archive interface design, calibration requirements, data accessibility and distribution, computing resources, and archive operations.

This Science Analysis Group [SAG #8] will analyze how the archive is to be used and scope the data requirements necessary to conduct science investigations related to the Cosmic Origins theme. The SAG will solicit input from the community to identify the types of investigations that will be conducted, and the kinds of data products that are valued and needed. The SAG will also consider what other assets or efforts may be needed to maximize the science return from the WFIRST archive (e.g., coordination of WFIRST-AFTA data with LSST, Euclid, or JWST; GO funding for ground-based observations or theoretical studies). The SAG will document its findings in a report to the Astrophysics Subcommittee.

The SAG8 charter can be conveniently divided into three parts. The first part describes and analyzes the WFIRST-AFTA Data Archive from the point of view of Cosmic-Origins scientists – what data, high-level science data products, and catalogs will be available in the WFIRST-AFTA archive; how to access the data; how well the data will be calibrated. The second part describes and analyzes the WFIRST-AFTA Data Archive from the point of view of computer scientists – archive interface design; computing resources; archive operations. The third part describes the WFIRST-AFTA Data Archive from a programmatic point of view – what coordination is needed with archives of other space or ground-based telescopes; funding for Guest Investigators of existing WFIRST-AFTA data, for Guest Observers obtaining new observations with WFIRST-AFTA, and for supporting theoretical studies.

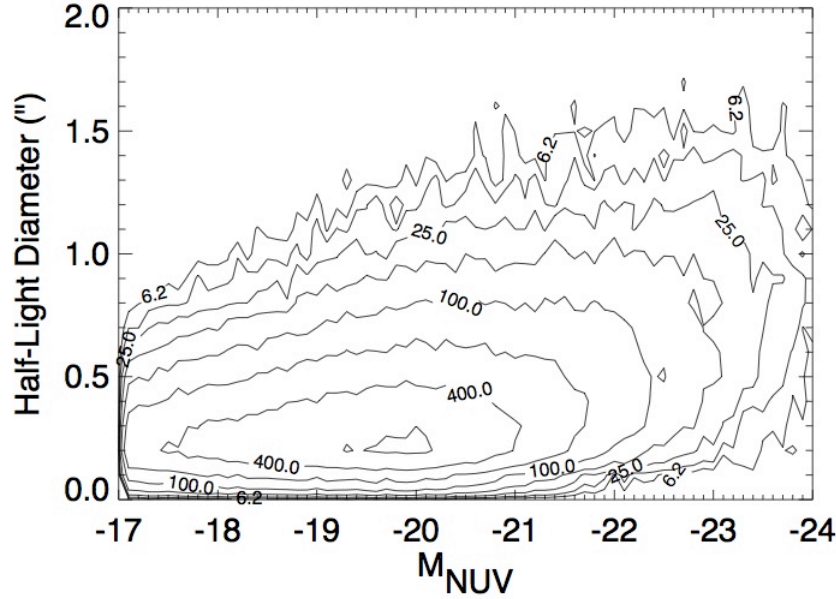
This report comes during an interim period between the Science Definition Team (SDT), which concluded its study with a report issued on March 10, 2015 (hereafter referred to as the “Report”) and a new Science Working Group whose make-up and responsibilities are as yet unknown. This report therefore only responds to the first part of the SAG8 Charter: an analysis of the WFIRST-AFTA Data Archive from the point of view of cosmic-origins scientists.

A Preview

The COSMOS Mock Catalog (CMC) (Jouvel et al. 2011) gives a preview of the WFIRST-AFTA survey. The CMC is a simulated catalog based on the COSMOS survey of 2 deg² by several observatories: GALEX for UV bands, Subaru for the optical (U to z), and CFHT, UKIRT, and Spitzer for the NIR bands. The CMC contains over 500,000 galaxies, AGN’s, and stars. The 111 items in the catalog give information such as type of galaxy, photometric redshift, E(B-V), half-light radius, AB magnitudes in various passbands ranging from Near-UV (Galex) to the IR (Spitzer IRAC). It also gives

predicted luminosities of major emission lines in the star-forming galaxies. The data are stored as a structure, so one can extract selected items from the catalog. The figure below gives an idea how the catalog might be used; in this case, to study the frequency of sizes and absolute near-UV magnitudes of galaxies at redshifts, $z=0.8-1.4$.

Figure 1: Plot based on data from the COSMOS Survey



WFIRST-AFTA will survey a high-latitude region 1000 times bigger than the COSMOS field; it will measure the shapes of 400 million galaxies as well as measure magnitudes in 4 passbands; it will obtain grism spectra of the full region. In addition, WFIRST-AFTA will survey the galactic bulge in search of micro-lensing events.

With WFIRST-AFTA's large and varied data sets, it will be impossible to troll a single file for the specific data desired. WFIRST-AFTA data need to be organized differently and stored redundantly for safety. There must be a way to find and access the data you want – and only the data you want -- efficiently.

Organization of Report

The following sections of this report are organized as follows:

Section 2 comprises excerpts from the *WFIRST-AFTA 2015 Report* describing Cosmic-Origins science (aka General Astrophysics) that can be accomplished with data obtained from the High-Latitude Surveys (HLS) and the Galactic Bulge Field.

The section also highlights Cosmic Origins science as described by potential Guest Investigators (GI's) using WFIRST-AFTA data and Guest Observers (GO's) who will make new observations of specific targets such as the Andromeda galaxy. Only a few examples are given in this report, but Appendix D of the *WFIRST-AFTA 2015 Report* reprints these 1-page science ideas in full.

In this shorter version of the report, Section 2 is omitted. Please refer to the WFIRST-AFTA 2015 Report for full details.

Section 3 briefly describes the data that will be obtained in each of the three High Latitude Surveys (Wide-Area Imaging for Weak Lensing, Wide-Area Spectroscopy for the Galaxy Redshift Survey and Galaxy clustering, the Supernova Field) and Wide-area imaging of the Galactic Bulge. It then goes on to describe the data processing of each of these types of observing programs and to give a (incomplete) list of on-line “catalogues” that will be produced.

Section 4 then describes how WFIRST-AFTA data, high-level science products, and catalogues can be found and accessed through querying the WFIRST-AFTA archive.

SECTION 2: GENERAL ASTROPHYSICS WITH WFIRST

The contents of this section are available in WFIRST SDT 2015 Report, which may be obtained from the WFIRST website:

http://wfirst.gsfc.nasa.gov/science/sdt_public/WFIRST-AFTA_SDT_Report_150310_Final.pdf

or from the arXiv:

Spergel et al., <http://arxiv.org/abs/1503.03757v2> .

Please refer to Section 2.3 of the 2015 Report, which describes general astrophysics with data from WFIRST High-Latitude Surveys.

Please refer to Section 2.5 of the 2015 Report, which describes general astrophysics with data from the WFIRST Galactic-Bulge Field.

Please refer to Appendix D of the 2015 Report, which contains submitted Guest Observer and Guest Investigator programs.

SECTION 3: WFIRST OBSERVATIONS, DATA PROCESSING, AND DATA PRODUCTS

As described in the SDT 2015 Report, WFIRST will make four major surveys including:

- The High Latitude Imaging Survey
- The High Latitude Spectroscopic Survey
- The Supernova Survey
- The Microlensing Survey

For each of these surveys, we give a description of the observations, required processing of the observational data, and the envisioned data products. The description of the observations given here is a lightly edited version of text in the 2015 Report, and the figures are screen captures of the figures in the 2015 Report, so the figure captions retain the same numbering system as the 2015 report.

The data description tables, data-processing, and data products from each survey have no counterpart in the 2015 Report. We have derived the needed data processing and data products based on knowledge of the scientific goals of WFIRST, WFIRST observational data, and previous experience with NASA's astrophysics space telescopes.

Introduction

The WFIRST-AFTA will make four major surveys:

- **The High-Latitude Imaging Survey** (HLS Imaging) will enable weak-lensing shape measurements of hundreds of millions of galaxies, which will in turn yield precise measurements of distances and matter clustering through measurements of cosmic shear, galaxy-galaxy lensing, and the abundance and mass profiles of galaxy clusters.
- **The High-Latitude Spectroscopic Survey** (HLS Spectroscopy) will measure redshifts of tens of millions of galaxies via slitless (grism) spectroscopy. Measurements of baryon acoustic oscillations (BAO) in these enormous 3-dimensional maps will pin down the cosmic distance scale and expansion rate at lookback times of 8 – 11 billion years (redshift $z = 1 - 3$), while measurements of redshift space distortions (RSD) will determine the growth rate of matter clustering over the same period.
- **The Supernova Survey** will combine wide-field imaging and Integral Field Unit (IFU) spectroscopy to measure precise distances to thousands of Type Ia supernovae (SNe). IFU observations will also be obtained in parallel with other HLS observations.
- **The Microlensing Survey** of the galactic bulge will search for microlensing events produced by planets in orbit around F, G, K, and M stars and beyond the “snow line”, complementing the exoplanets found by Kepler, which orbit closer to the star.

The four surveys will be scheduled according to a draft Design Reference Mission shown in the figure below (source: SDT 2015 Report, p. 139).

Graphical Design Reference Mission

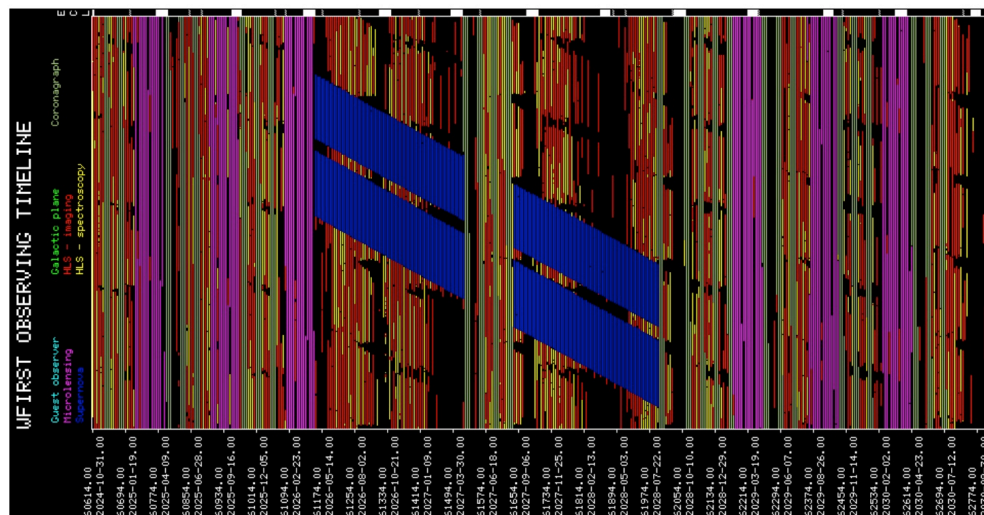


Figure 3-53: The example WFIRST-AFTA observing sequence. Each vertical stripe denotes a 5-day period, and the 6-year mission proceeds from left to right. The microlensing seasons (magenta) and coronagraph blocks (gray) are indicated, as are eclipse seasons (white rectangles at top). The supernova survey (blue) uses portions of the telescope time for 2 years. High-latitude imaging (red) and spectroscopy (yellow) are interspersed.

3.5 Calibrations. The four major surveys described above will not only yield important scientific information directly but will be used to produce essential calibrations. These include:

- Calibration of photometric redshifts with spectroscopic redshifts
- Astrometric fiducials for high-latitude stars
- Characterization of the Wide-Field Imager (WFI) sensor array

Scope of this Report

As described in the 2015 Report, all four surveys will produce important information for cosmic-origins science, *without modification to the observing program*. This section describes the observational data WFIRST will obtain in these surveys and the processing they will undergo *in order to be useful for Cosmic Origins-related studies*. We expect that in most cases the data processing for Cosmic Origins science will coincide with that needed for dark energy and microlensing studies. Nevertheless, this report should not be construed as a prescription for data processing by the survey teams, who know best how to treat the data for their own purposes. This report speaks only on behalf of Cosmic Origins scientists.

Another important caveat is that the WFIRST-AFTA SDT was not charged with developing a plan for data processing, archival and distribution. Nor are the detailed techniques for treating the data a settled matter. For example, it is not clear at this point whether instrumental distortions will be removed before or after galaxy shape measurements proceed.

Data Processing Levels

NASA customarily classifies science data according the level of processing received. The data-processing steps are described in general terms. Here, we describe the data processing steps as applied to WFIRST-AFTA data. Typically (but not always), the WFIRST science data-processing system:

- *Acquires science data* (images, spectrograms) and engineering (e.g. temperatures) in the telemetry data and corrects for/removes telemetry artifacts and extraneous data (Level 0 data).

At this early stage of WFIRST-AFTA, the contents of the science and engineering data down-linked to the ground are uncertain because the orbit of WFIRST-AFTA has not been selected. If the orbit is geo-synchronous then all the data read out from the Wide Field Imager (WFI) every 5 sec or so can be sent to the ground, but if the orbit is about the L2 libration point, only selected data – say, the counts accumulated after 6 non-destructive readouts (30 s) – can be sent to the ground. In the latter case, the data would be processed on-board and

corrected for cosmic-ray artifacts according to an algorithm such as Offenberg, Fixsen, & Mather (2004). However, the selected orbit and on-board processing are not settled matters at this time.

- *Gathers individual exposures* into an observation set (multiple observations at a nominal pointing but different dithers), resamples, and reformats data into a single image or spectrogram in FITS format, and in the process, corrects for cosmic-ray and instrumental artifacts (Level 1 data);
- *Removes the instrument signature* using calibration data and algorithms to yield calibrated data (Level 2 data). Both photometric calibration and geometric calibration (correction for instrumental optical distortions) are carried out. Also wavelengths are assigned to grism spectra or IFU spectra.
- *Operates on and makes measurements on the calibrated data* to produce a Source List with measured properties (Level 3 data). Each observation set produces a Source List.
- *Associates each Source with an Object* in various Object Lists, which contain or have pointers to all measured properties, spectra, postage-stamp images, etc. of each object observed by WFIRST-AFTA (Level 4a data).
- *Combines/resolves differences in the measured properties of an Object* derived from different WFIRST-AFTA observations and possibly other observatories (Level 4b data).
- *Produces “catalogues” of objects* (in reality, on-line relational databases, maps, etc.) of use to the astronomical community (level 4c data). Some examples:
 - Catalog of galaxy spectroscopic redshifts resulting from IFU observations
 - Catalog of galaxy photometric redshifts derived from LSST and WFIRST-AFTA filter fluxes
 - Catalog of proper motions of stars in the galactic bulge derived from numerous astrometric measurements over the course of the observing program
 - Catalog of SNe, with classifications, redshifts, etc. and association with host galaxies

The Object Catalog will be extended as new observations arrive and revised as new calibration algorithms and data become available. Since most of the surveys will take the full primary mission (6 years) to complete, the Object Catalog at any point in time should be considered a work in progress. If WFIRST-AFTA follows the example of SDSS and LSST, the full data set will be reprocessed at least annually. This means full reprocessing of the observed data from Level 1 on.

High-Latitude Imaging Survey (HLS Imaging)

Over its six-year primary mission, the Wide-Field Imager (WFI) will map $\sim 2,200 \text{ deg}^2$ of sky in four broad NIR passbands (Y, J, H, F184) down to a 5-sigma limiting AB magnitude of $J=26.9$ (point source). The WFI field of view is 0.281 deg^2 . The image is recorded by 18 4k x 4k HgCdTe sensor arrays (288 million pixels; ~ 0.6 Gigabytes) (Figure 1-1 on next page). Each square pixel subtends 0.11 arcsec. The exposure time for each of the filters is 174 s at each dither position. The HLS imaging survey follows a dithering strategy (Fig 3-54 on next page) so that images in J, H, and F184 are fully sampled for shape measurements, even when some exposures of a galaxy are lost to chip gaps, cosmic rays, or detector array defects. Thus a given galaxy is observed 8-10 times in the J filter, 5-8 times in Y, H, and F184. (J-band images get extra dithers for improved PSF sampling.)

The total exposure time for the imaging survey is 2.01 years including overheads. The WFI observations are expected to detect, identify, and measure the shapes of 380 million galaxies. The shape measurements are based on observations in 3 filters, 2 roll orientations per filter, and 3-5 dither positions per orientation. The WFI observations are also expected to lead to the discovery of $>10,000$ high-redshift ($z>8$) galaxies.

Processing of Wide-Field Images For Cosmic Origins Science

Data Level	Data Description
Level 0	Single-frame, reformatted, unprocessed science image(s) and ancillary data with telemetry artifacts removed
Level 1	Resampled image(s) derived from level-0 images at other dither positions and corrected for cosmic-ray artifacts
Level 2	Level-1 image with geometric, photometric, and astrometric calibrations applied given the best calibration algorithms & data available at the time
Level 3	Measurements of Level-2 images to produce a Source List with measured properties such as source id, source type, filter ID, AB magnitude, position, size and shape
Level 4	Creation of or addition to the HLS Imaging Object List by association of each Source to an Object

Science Data Products: Indexed Relational Databases & Lists - Examples

- Stellar proper motions of the Sagittarius stream
- Photometric & astrometric (parallax, proper motion) properties of halo stars
- Catalog of rich galaxy clusters
- Overdense regions in image (star clusters or streams, galaxy clusters)
- Catalog of very high-redshift ($z_{\text{phot}}>7$) galaxies
- List of astrometric fiducials (barely resolved galaxies) for this pointing

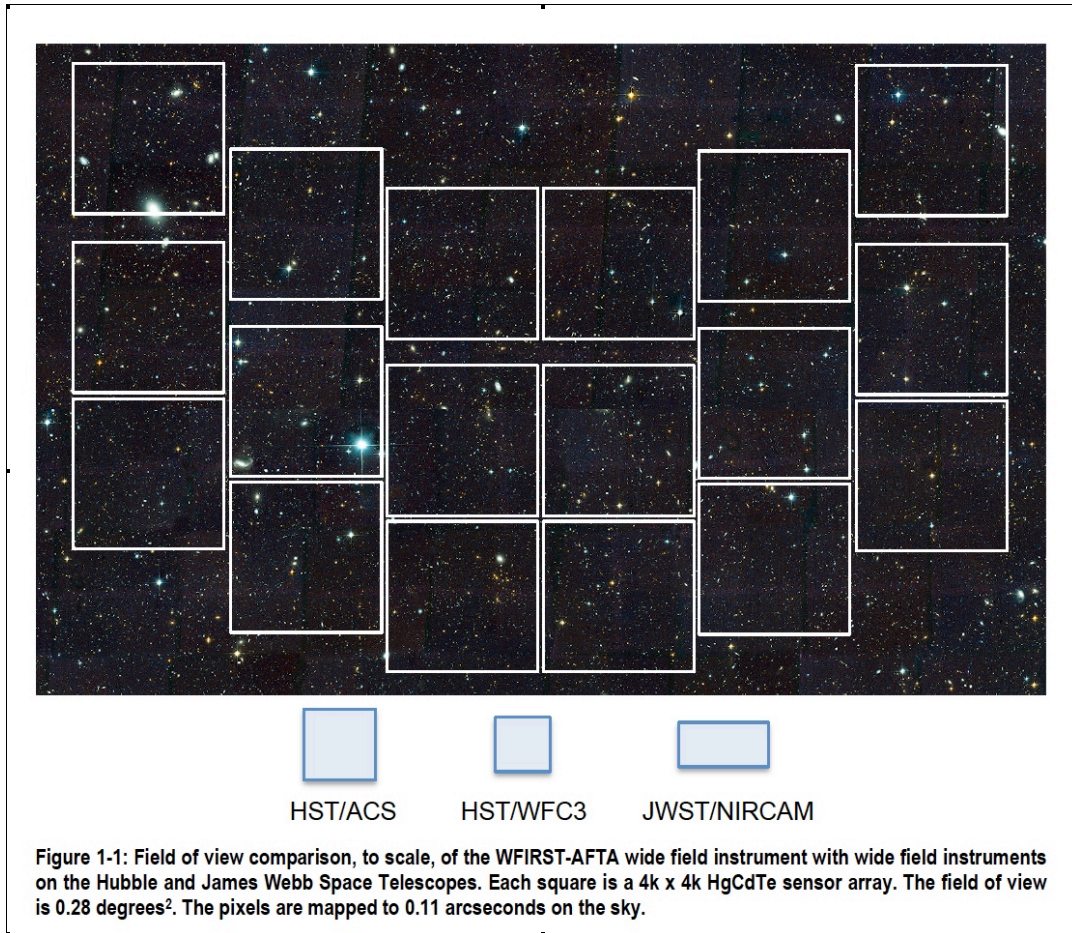
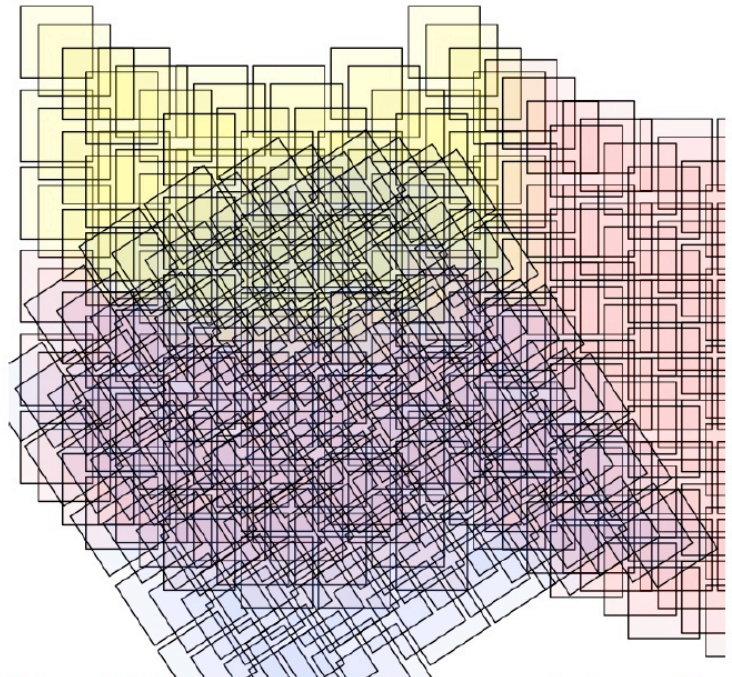


Figure 3-54: Example of the sequence of observations for the high-latitude survey in H-band. The observatory first performs a sequence of 4 small-step dithers to cover the chip gaps (yellow). This pattern is repeated to cover the sky (red). A half-integer number of years later, the observatory returns to perform a second pass over the field (blue) at a general angle. The second pass enables internal relative calibration via repeat observations of stars at all pairs of positions on the focal plane, as well as monitoring of long-term drifts.



High Latitude Spectroscopic Survey

The High Latitude Spectroscopic Survey will make use of a grism to obtain slitless spectrograms over the same region of the sky as the HLS imaging map (Fig. 2-16 below). The exposure time for each observation (dither position, roll angle) will be 347 s. The observation will be repeated after a small dither, and another pair of pointings will be obtained after a roll of 4-5 degrees. This observation set will be repeated at ≥ 2 different telescope roll angles, two of which are approximately opposed. When co-added together, the R~600 spectra covering $\lambda=1.35\text{-}1.95\text{ }\mu\text{m}$ will have a 7-sigma line-flux sensitivity of $0.5 \times 10^{-16}\text{ erg/s/cm}^2$ (point source). The survey is expected to detect and identify 16 million Ha-emitting galaxies at $z=1.06\text{-}1.88$, and 1.4 million [O III]-emitting galaxies at $z=1.88\text{-}2.77$. It will also obtain spectra of non-emission-line objects.

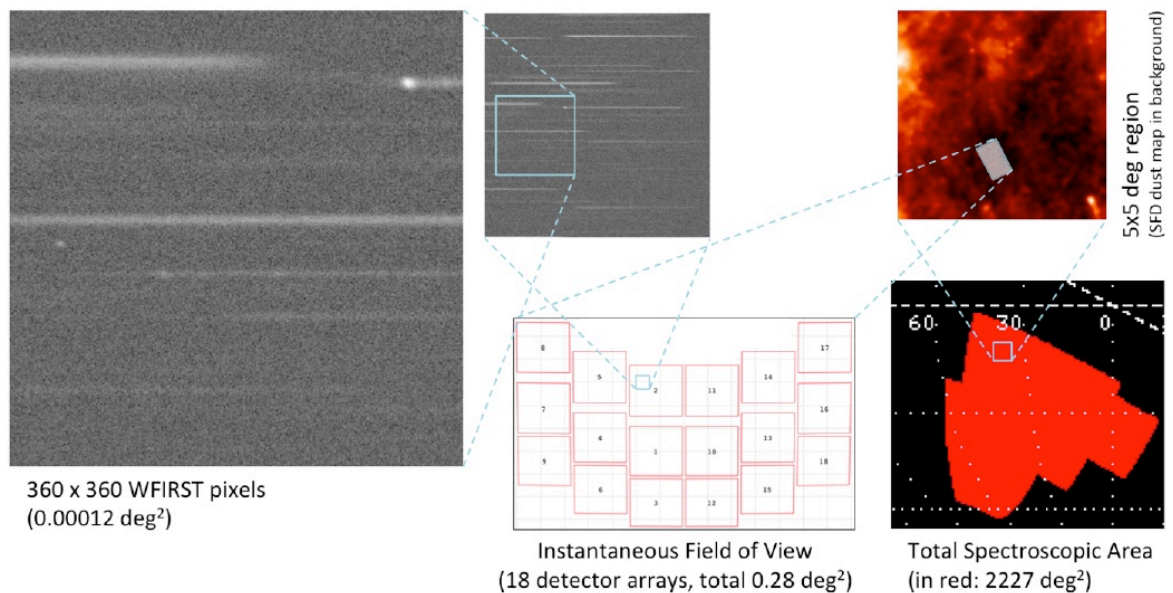


Figure 2-16: A simulated 1st order grism image at WFIRST-AFTA resolution and sensitivity, and at 347 s exposure time as planned for the HLS. Each field in the HLS will be covered with a median of 7 grism exposures at 4 roll angles. This image covers 40x40 arcsec; the instantaneous field of view of WFIRST-AFTA is 2300 times larger.

Processing of Grism Spectrograms for Cosmic Origins Science

The full set of grism data at a particular location on the sky consists of 4 pairs of spectrograms: one pair at the initial roll angle, one pair offset by 4-5 degrees, and two pairs offset from the first by 160° - 180° . Thus, the basic unit of grism data is a pair of spectrograms at 2 dither positions but the same telescope roll angle.

Data Level	Data Description
Level 0	Two unprocessed spectrograms and ancillary data with telemetry artifacts removed
Level 1	Spectrogram combined from the two dithers and corrected for cosmic-ray artifacts
Level 2	Extracted spectra of stars and galaxies with wavelength and flux calibrations applied given the best calibration algorithms & data available at the time
Level 3	Creation of or addition to Source List including measured galaxy ID, wavelengths and fluxes (erg/s/cm^2) of identified spectral lines, estimated spectroscopic redshift, and pointers to the calibrated spectra.
Level 4	Co-addition of spectra of each Source at all dithers and telescope roll angles and association of each Source to an Object in the Object List.

Science Data Products: Indexed Relational Databases & Lists (Examples)

- Galaxy Redshift database
- Ha-emitting galaxies - properties
- Luminous red galaxies- properties
- High-redshift galaxies/AGN's - properties

IFU Survey of Supernovae and Galaxies

The Supernova Survey field is a 27.44 deg^2 field (~ 100 WFI fields) consisting of three nested tiers: a shallow survey for $z < 0.4$ SNe covers the whole field; a medium-deep survey, over 8.96 deg^2 for $z < 0.8$ SNe; and a deep survey, over 5.04 deg^2 for SNe out to $z = 1.7$. This field will be surveyed in just over six months total observing time spread over 2 years. It is expected that 2725 SN Ia's will be discovered.

Before the survey begins in earnest, the field will be observed with multi-filter imaging with the Wide Field Imager (WFI) (and, ideally with spectra from the Wide Field Grism, as well) in order to obtain photometric redshifts (and where possible, spectroscopic redshifts) for most of the galaxies that will later host supernovae.

Promising SN Ia candidates will be observed with the Integral Field Unit (IFU) every 5 days (in the rest frame of the host galaxy). The IFU channel uses an image slicer and spectrograph to provide individual $R \sim 100$ spectra of each slice covering the $0.6 - 2.0 \text{ }\mu\text{m}$ spectral range over two small fields: a $3.00 \times 3.15 \text{ arcsec}$ field with $0.15''$ slices, and a $6.00 \times 6.30 \text{ arcsec}$ field with $0.30''$ slices. Both fields are imaged onto a single $2\text{k} \times 2\text{k}$ HgCdTe detector.

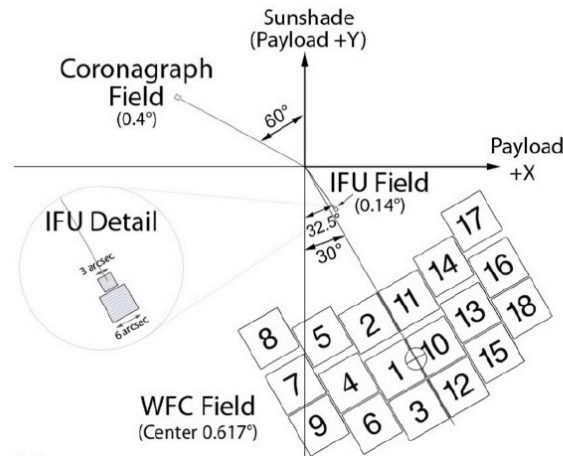


Figure 3-1: The WFIRST-AFTA fields of view layout as projected on the sky showing the wide-field instrument fields (wide-field and IFU channels) and the coronagraph field (shared between the imager and the IFS).

The IFU will also observe small fields in parallel with high-latitude surveys with the Wide-Field Imager (WFI) and Wide-Field Spectrograph (WFS). Keeping step with the WFI or WFS means that the IFU exposure time is also about 173 s and that the IFU will sample up to 10 small fields for every wide-field observation. As the WFI or WFS returns to the original field, the IFU fields will be re-observed. In this way, the IFU is expected to obtain spectra covering $0.6\text{--}2.0 \text{ }\mu\text{m}$ of 20,000 – 30,000 galaxies, thereby providing spectroscopic redshifts for calibrating photometric redshifts.

Processing of IFU Spectra for Cosmic Origins Science

Data Level	Data Description
Level 0	Unprocessed spectrogram with telemetry artifacts removed
Level 1	Level-0 data (spectrogram) corrected for cosmic rays and reformatted into a datacube (x,y,λ) , so that a x-y slice of the datacube produces a monochromatic image at a particular wavelength, and a drill through the datacube at a particular x,y yields a R~100 spectrum at that location.
Level 2	Extracted spectra of galaxies in the IFU field of view with wavelength and flux calibrations applied
Level 3	<p>Measurements of the extracted spectra to yield spectroscopic redshifts, emission-line ID's, etc. and creation of or addition to a Source List with pointers to the calibrated spectrum and measured fluxes</p> <p>For SN Ia candidates: Measurement of the SED to yield the flux in selected wavelength bands for that epoch for insertion into a table of light curves for that source.</p>
Level 4	<p>Association of each Source with an Object in an Object list with spectra, spectroscopic redshift, and light-curve fluxes for that epoch (level 4a data).</p> <p>Combination of spectra of each source at different dithers and roll angles (level 4b data)</p> <p>Production of databases containing spectra, spectroscopic redshift, and imaging data.</p> <p>Production of light-curve catalog at all epochs (so far).</p>

Science Data Products: Indexed Relational Databases & Lists - Examples

IFU spectroscopic redshift catalog
 SN Ia Light Curves at various wavelengths
 SN Ibc's and other variable sources in the Supernova field
 SN Ia Host Galaxies - properties & spectra

Microlensing Survey of the Galactic Bulge

The Wide Field Imager will also be used to make a microlensing survey of the Galactic Bulge. The survey will cover ten contiguous fields between Galactic longitudes ~ -0.4 to 1.8 deg, and Galactic latitudes ~ -1 to 2.2 deg, for a total area of 2.81 deg^2 . The ten fields will be imaged in six 72-day campaigns spread over the nominal six-year mission (Fig 3-53). In each campaign, each field will be observed every 15 minute in a wide (W149) filter in a 52-sec exposure, and once every 12 hours in a bluer filter (Z087) in a 290-sec exposure. In 85% of the survey footprint, the total number of epochs will be $\sim 33,000$ in the W149 filter, and ~ 700 in the Z087 filter. The strategy on dithers and addition of a bluer filter for the Wide Field Imager are under discussion.

Among the 10^8 stars viewed in the microlensing survey, ~ 2600 bound exoplanets are expected to be detected in the range, 0.1 - $10,000$ Earth masses including ~ 370 Earth-mass planets, and ~ 30 free-floating Earth-mass planets if there is one per star in the Galaxy. It will also discover $\sim 20,000$ transiting planets with orbits less than a few tenths of an AU.

The microlensing survey is also expected to:

- Measure or constrain the mass function of compact objects over roughly 8 orders of magnitude, from isolated compact objects with masses from that of Mars to roughly 30 times the mass of the Sun;
- Detect and measure the mass of over 100 neutron stars and black holes;
- Yield a very deep color-magnitude diagram of the $\sim 3 \text{ deg}^2$ microlensing field;
- Measure the luminosity function of stars in the bulge down to nearly the bottom of the main sequence, as well as to identify unusual stellar populations;
- Identify many variable sources with amplitudes $> \text{few mmag}$, timescales from tens of minutes up to the ~ 6 -year duration of the mission;
- Determine the mass and radii of $\sim 1 \times 10^6$ giant stars brighter than $H_{AB} = 14$ stars through asteroseismology;
- Obtain photon-noise-limited parallaxes of $< 10\%$ and proper motion measurements of $< 0.3\%$ (0.01 mas/yr) for the ~ 56 million bulge and disk stars with $H_{AB} < 21.6$ in its field of view.

If combined with photometry in a filter bluer than Z087, the microlensing survey will also yield:

- An estimate of the effective temperature, metallicity, age, luminosity, and foreground extinction for all stars;
- A determination of the bulge mass and velocity distribution (including bar structure);
- The stellar density and velocity distribution of the Galactic disk;
- The metallicity and age distribution of the disk and bulge, and the three-dimensional distribution of dust along the line of sight toward the bulge fields.

Processing of WFI Survey of the Galactic Bulge for Cosmic Origins Science

Data Level	Data Description
Level 0	Reformatted, unprocessed wide-field image with telemetry artifacts removed
Level 1	Level-0 image combined with other dither positions and corrected for cosmic-ray artifacts Image difference (image subtracted from level-1 previous image)
Level 2	Level 1 data with geometric, photometric, and astrometric calibrations applied given the best calibration algorithms & data available at the time
Level 3	Creation of or addition to Bulge Source List with measured properties e.g. source id, time, filter ID, AB magnitude, position, etc.
Level 4	Creation of or addition to the Bulge Object List by association of each source to an object observed at a different epoch

Science Data Products: Indexed Relational Databases & Lists - Examples

- Catalog of micro-lensing events
- Catalog of massive stellar remnants (black holes and neutron stars) in the bulge field
- Compact objects with masses ranging from Mars-like masses to 30 Msun
- Monitor of the WFI detector characteristics (see 2015 Report 2.5.7)
- Catalog of stellar populations in the Galactic Bulge including measurements of position, AB magnitudes at W149 and Z087, color (Z087-W149), proper motion, parallax

Special Calibrations

Calibration of Photometric Redshifts (2015 Report, §2.2.3.4)

Photometric redshifts for a sample of hundreds of millions of galaxies out to $z=3$ will be a huge asset to studies of the evolution and clustering of galaxies and AGN. Photometric redshifts (z_{phot}) of galaxies in the HLS imaging survey will rely on a combination of WFIRST-AFTA data and optical data from the ground, e.g. LSST. The SDT's simulations indicate that the combination of WFIRST+LSST provides excellent performance out to $z>3$. The estimated rms uncertainty of z_{phot} is $\sigma_z/(1+z)=0.061$, and estimated fraction of objects with fractional redshift error < 0.04 is 78%.

One strategy for improving the photometric redshifts involves calibration with spectroscopic redshifts obtained with the WFIRST-AFTA IFU. The IFU will observe in parallel with the WFI during its weak-lensing survey, sampling numerous fields of galaxies and yielding an estimated $\sim 20,000$ - $30,000$ spectroscopic redshifts for these galaxies.

Astrometric Fiducials for Measuring the Proper Motions of High-Latitude Stars (2015 Report §2.3.3)

WFIRST-AFTA will provide access to an enormous number of slightly resolved medium-brightness galaxies, and absolute motions can be measured directly within the field of each detector array. This local approach to measuring absolute motions has recently been used in the optical with HST to measure the absolute motions of the globular clusters, LMC, SMC, dwarf spheroidals and even M31, in addition to individual hyper-velocity and field stars. The strategy involves constructing a template for each galaxy so that a consistent position can be measured for it in exposures taken at different epochs. This template can be convolved with the PSF to account for any variations in focus or changes of the PSF with location on the detector array. The same approach that worked with HST in the optical should work with WFIRST-AFTA in the IR. [Based on prior experience with HST], the SDT expects about 500 reference objects in each WFIRST-AFTA detector array, which can be used to tie down the absolute frame to better than 0.5 mas in each exposure. The HLS 2-year baseline would allow absolute motions to be derived with systematic accuracies of about $125 \mu\text{as}/\text{year}$. Follow-on GO programs could extend this baseline to 5-years, enabling accuracies of about $50 \mu\text{as}/\text{year}$. There are about 10 stars in the ~ 5 square arc-minute FoV of the UDF that can be measured with this accuracy, implying about 18 million halo stars in the high latitude imaging survey.

Wide-Field Imager Detector Characterization (2015 Report §2.5.7)

With its $\sim 33,000$ dithered images of $\sim 10^8$ point sources with known colors and nearly constant fluxes, the Microlensing Survey of the galactic bulge will likely provide one of the best datasets with which to characterize WFIRST-AFTA's wide-field imager. In particular, we imagine that the bulge survey pipeline will measure the fluxes and positions of all of the stars in the field of view, while simultaneously measuring the response function of each of the WFI pixels, and how this response function varies with time. This time-variable response function will then be used to calibrate and remove detector artifacts such as persistence, inter-pixel capacitance, reciprocity failure, non-linearity, and intrapixel response variations.

SECTION 4. ACCESSING THE WFIRST-AFTA ARCHIVES

This section focuses on the means by which scientists will identify, access, and extract information from the archives. It recommends a query system as the main interface between the user and the archives. To illustrate how this would work, we give examples of queries that return the most relevant information and can be used to discover complex relationships. The queries given here were inspired (not exact quotes) by:

- the 1-page science ideas, printed in Appendix D of the *WFIRST-AFTA SDT 2015 Report*, March 10, 2015

http://wfirst.gsfc.nasa.gov/science/sdt_public/WFIRST-AFTA_SDT_Report_150310_Final.pdf

- A. Szalay et al.'s 20 Questions for the SDSS Skyserver

Szalay et al. (2000), *Designing & Mining Multi-TB Ast. Archives*, Proc. ACM SIGMOD 2000, p. 451

- The *WFIRST-AFTA SDT Interim Report*, Apr 30, 2014

http://wfirst.gsfc.nasa.gov/science/sdt_public/WFIRSTAFTA_SDT_Interim_Report_April_2014.pdf

- J. Kruk (private communication)

The reference for each query is given in shorthand, e.g. (D-12, Tanner) for A. Tanner's 1-page science idea printed in Appendix D, page 12; (Szalay, Q1) for Question #1 of Szalay's 20 Questions for the SDSS Skyserver; (Gould, IR-57) for the *SDT Interim Report*, pg. 57).

Introduction

A single high-latitude imaging observation will record light from nearby objects (e.g. asteroids, brown dwarfs, cool dwarf stars), halo stars of the Milky Way, stellar streams orbiting the Milky Way, nearby galaxies, and distant galaxies. All these objects are a delight to cosmic-origins scientists.

As described in Section 3, the WFIRST archive will consist of two types of data. Level 0-2 science and calibration images will be stored in “flat” files, presumably in FITS. For example, Wide Field Imager or Spectrograph data will be stored in 18 FITS extensions -- one extension for each of the 18 detectors. Similarly, IFU data cubes may be stored in 20 FITS extensions – one extension for each slice. In the case of spectra extracted from calibrated grism images or IFU data cubes, the data may be stored in FITS structures with each object having 1-D data (wavelength, flux, propagated error, quality flag, etc.) In contrast, most Level 3+ data will be stored in numerous, heavily indexed, relational databases (often called “catalogues” for brevity in this report). These databases contain the results of measurements on the data, source lists, and object lists, etc.

Most data of interest to cosmic-origins scientists will be contained in one or more catalogues rather than the observational-data files from which they were derived. For example, the study of kinematics of stellar streams orbiting the Milky Way (2015 Report §2.3.3 reprinted in Section 2 of this report) will identify stellar streams by their common proper motion as listed in the HLS/WFI Stellar Proper Motion database.

Cosmic-origins scientists will rely on sophisticated Level 3+ data, beyond the most common measurements of position, photometry and line strength. They will expect that the WFIRST archive will contain the output of significant data processing efforts by the science centers and the survey teams. Quantities such statistical clustering parameters or advanced morphological classifications (galaxy types, lenses, tidal features, etc.) may be required. We expect that advances in computing and software technology, such as machine learning, will enable tremendous advances in what the archive can offer.

Beyond even the advanced data products that the archive can offer, cosmic-origins science will also rely on the capability of the archive perform new kinds of analysis “near the data”. Given the sheer volume of the WFIRST and ancillary data sets, it may not be practical for researchers to do all of the analysis at home. Rather, the archive will likely need to set up the infrastructure to accept (reasonably sized and vetted) software to be run on CPUs with immediate access to the “big data” stored at the archive. This new paradigm will help level the playing field between the few institutions that can afford to build major computing centers and others where good ideas would otherwise be put aside.

Given the complexity of WFIRST (and ancillary) data, providing easy access will be a challenge on its own. Fortunately, Cosmic-Origins scientists don’t need to know where the relevant data are stored; they can get the information they need unencumbered by

irrelevant data by querying the WFIRST-AFTA archive. The SDSS archives and NASA archives like WISE and 2MASS at IRSA and the Hubble Source Catalog at MAST provide a query system to help a user find the objects of interest. For example, a user might request the database to return the positions of all stars in the Small Magellanic Cloud brighter than some magnitude redder than some J-H color.

The archives use Structured Query Language (SQL) for all queries. These queries can range from the simple to highly complex. The figure below shows examples of queries to the SDSS archives. Users can modify a sample query similar to their own without needing detailed knowledge of SQL. The set of sample queries should be useful for most queries of the WFIRST-AFTA archive.

Basic SQL: Basic SELECT-FROM-WHERE Basic position search Using PhotoTag Search for a Range of Values Rectangular position search More than one table: JOIN...ON Photometry & spectroscopy Counting by type or category Using flags	General Astronomy: Only stars or galaxies Clean photometry Using Field MJD Objects by spectral lines Spectra by classification Moving asteroids Plates with repeat spectra Galaxies blended with stars Counts by type and program Checking SDSS footprint	Galaxies: Clean photometry - Galaxies Galaxies with blue centers Diameter limited sample LRG sample selection Galaxy counts on HTM grid Classifications from Galaxy Zoo BOSS target selection BOSS Stellar Masses BOSS Stellar Vel. Disps.
SQL Jujitsu: Data subsample Objects in close pairs Selected neighbors in run Object counts and logic Repeated high-z objects Splitting 64-bit values Using LEFT OUTER JOIN Using Nested Queries	Stars: Clean photometry - Stars CVs using colors Binary stars colors Using sppLines table Using sppParams table Proper motions	Variability Queries: Stars multiply measured Multiple Detections and Time Series
Quasars: QSOs by spectroscopy QSOs by colors FIRST matches for quasars	Miscellaneous: Photometric Redshifts Spectra in Other Programs - I Spectra in Other Programs - II Using WISE Cross-Match	APOGEE: All APOGEE Plate Visits ASPCAP Parameters and Errors APOGEE Stars No BAD Flags ASPCAP Params for Cluster Mbrs APOGEE Proper Motions APOGEE Stars Near Cluster Ctr RVs for Individual APOGEE Visits APOGEE and SEGUE Spectra SDSS photometry for APOGEE Stars

Source: <http://skyserver.sdss.org/dr10/en/help/docs/realquery.aspx>

The remainder of this section gives sample questions specific to WFIRST-AFTA observations that can be translated into SQL queries, if the relevant products have been produced by the data processing efforts. These questions can be used to perform stress tests on the WFIRST-AFTA Archive while in development. The questions are inspired (not exact quotes) by the 1-page science ideas (printed in Appendix D of the WFIRST-AFTA 2015 Report), by Szalay et al.'s (2000) *20 Questions for the SDSS Skyserver*, by Gould in an *WFIRST Intermediate Report* (IR), and by J. Kruk.

Sample Queries of the WFIRST Archive

A. MICROLENSING SURVEY (using filters Z087, W149)

ML1: Find all microlensing events of stars in the galactic bulge in which the apparent position of the lens shifted by more than $\langle X \rangle$ micro-arcsec during the microlensing event. (Sahu, D-18). This is a search for neutron stars and stellar-mass black holes in the Galaxy.

ML2: Provide a list of all bulge stars showing evidence of having a transiting planet(s). (Gould IR-53,57)

ML3: Provide a list of all KBO's detected by WFIRST (Gould 2014, IR-57)

ML4: Find all objects whose absolute magnitudes and colors are consistent with blue stragglers /red giants /young, massive stars/ white dwarfs / <keyword>. (31Oct14 WFIRST SDT telecon)

B. HIGH-LATITUDE IMAGING SURVEY (using filters Z087, Y, J, H F184)

Milky Way and Local Group Science

HLIL-1: Find all stars brighter than $J \sim 25$ whose LSST + WFIRST + WISE colors are consistent with an L or T brown dwarf. (D-12, Tanner)

HLIL-2: Find statistically overdense regions of resolved stars consistent with dwarf-galaxy satellites of the Milky Way. (D-22, Geha)

HLIL-3: Find all stars in a selected nearby galaxy ($d < 5$ Mpc) with absolute magnitudes and colors consistent with red-giant stars. (D-27, van der Marel)

HLIL-4: Find all galaxies within 50 Mpc showing evidence of tidal stellar streams or intrachuster light. (D-29, Laine)

HLIL-5: Provide a list of stars that are 1% rare for LSST+WFIRST-color attributes. (Szalay, Q7)

HLIL-6: Provide registered 4-filter images of a selected star-forming region that can be used to create a dust map of the region (IR-71)

HLIL-7: Provide a map of absolute proper motions of stars in the Small Magellanic Cloud (2015 Report §2.3.3).

Science Beyond the Local Group

HLIG-1: Find all galaxies within $\langle x \rangle$ arcsec of a given point in the sky (Szalay, Q1) This is a classic spatial lookup.

HLIG-2: Find all objects within $1'$ of one another that have similar colors. This is a gravitational lens query. (Szalay, Q18)

HLIG-3: Find all objects with rings or arcs. This is a search for strongly lensed galaxies (Stern, D-33; Fan, D-43; IR-48)

HLIG-4: Find all lensed galaxies images brighter than $J=X$ (bright enough for follow-up spectroscopy by large ground-based telescopes).

HLIG-5: Find all galaxies with central surface brightness fainter than $J = 23$ mag per square arcseconds. (Szalay, Q2)

HLIG-6: Find all galaxies showing a tidal feature with no obvious companion in the H band (Conselice, D-32)

HLIG-7: Find galaxies that are blended with a star; output the deblended magnitudes. (Szalay, Q6)

HLIG-8: For a selected massive galaxy cluster, find all galaxies within $30''$ of the brightest cluster galaxy (BCG) with a photo- z within 0.05 of the BCG. (Szalay, Q20)

HLIG-9: Find all galaxies satisfying user-specified color cuts and magnitude & redshift ranges (Szalay, Q13).

HLIG-10: Find all $z > 8$ galaxies redder than $(H-F184) = \langle X \rangle$

HLIG-11: Find all regions of the sky where the surface density of galaxies observed in the H-band is greater than 50 galaxies per square arc minute (search for galaxy clusters, 2015 Report §2.3.2)

HLIG-12: Find all galaxies brighter than $J=26$ with shape measurements and 4-filter $z_{\text{phot}} > 7.5$ (search for lensed high- z galaxies, 2015 Report §2.3.1)

C. HIGH LATITUDE SPECTROSCOPIC SURVEY (λ range=1.35-1.95 μm)

NOTE: Some of these queries require data from not only WFIRST (YJHF184) imaging and spectroscopic catalogs ($\lambda\lambda$ 1.35-1.95 μm) but also the LSST (*ugrizy*) catalogs and the Euclid spectroscopic catalog ($\lambda\lambda$ 1.25-1.85 μm).

HLS-1: Provide a histogram of H-alpha fluxes and equivalent widths as a function of spectroscopic redshift in the interval, $z \sim 1-2$. Do the same for [O II] 3727 in the redshift interval, 2.6-4.3. (2015 Report §2.3.4)

HLS-2: Provide a plot showing IFU z_{spec} 's vs. LSST+WFIRST z_{phot} 's (Report 2.2.3.4)

HLS-3: Provide a color-luminosity plot of weak-lensing galaxies in a selected narrow redshift bin with symbols/contours indicating the size of (zphot-zspec) error (Report 2.2.3.4)

HLS-4: Find spectra of all Lyman-a emitters (LAE's) at $z > 7$.

HLS-5: Find all quasars at $z = 8-10$ (D-44, Fan)

HLS-6: Provide spectroscopic redshifts for LSST objects of interest identified by their RA, Dec, and object type. (D-48, Strauss)

HLS-7: Provide IR SED's for LSST stars of interest (D-48, Strauss)

HLS-8: Find all elliptical galaxies at $z > 1$ whose spectra show an anomalous emission line (Szalay, Q11)

HLS-9: Find all objects with spectra unclassified. (Szalay, Q8)

HLS-10: Find quasars with a broad absorption line in their spectra and at least one galaxy within $10''$. (Szalay, Qxx)

HLS-11: Find quasars with a line width > 2000 km/s and $2.5 < \text{redshift} < 2.7$. (Szalay, Q9)

HLS-12: Find galaxies with spectra that have an equivalent width in $H\alpha > 40\text{\AA}$ (Szalay, Q10)

E. TIME-DOMAIN INFORMATION

TD-1: Find all the objects in the microlensing survey exhibiting peaks in their lightcurve with significance greater than 5-sigma and with durations shorter than 10 days. (Kruk)

TD-2: Download all observations encompassing a given spot on the sky within $\pm <X>$ days of a given date. (Kruk)

TD-3: Find all objects whose J-band fluxes have changed by more than $<X\%>$ since measured by HST. (Kruk)

TD-4: Find all galaxies in clusters at $z \sim 0.5-1.5$ in which > 0.5 -mag variability was detected at F184W (Kruk)

F. IFU SURVEY

IFU-1: Provide a list of all red galaxies showing $H\alpha$ in emission.

IFU-2: Download the light curves, redshifts, and imaging data of all SN Ia's

IFU-3: Provide the AB magnitudes, sizes and shapes of all host galaxies of SN Ia's

IFU-4: Provide a list of galaxies showing variability not associated with an SN Ia

IFU-5: Provide postage-stamp images of all galaxies hosting a SN Ia, both with the bright SN and after the SN has faded

G. GENERAL OBSERVER OBSERVATIONS

GO-1: Provide a list of globular clusters / nearby galaxies / galaxy clusters / <keyword> observed by WFIRST

GO-2: Download the complete Object List containing all associated measurements for the WFIRST Deep Field